Beyond Prototypes: Challenges in Deploying XR Smart Glasses as Assistive Devices for People with Cerebral Visual Impairment

Anonymous ICCV submission

Paper ID *****

Abstract

Cerebral Visual Impairment (CVI) is set to become the leading cause of vision impairment but remains underrepresented in assistive technology research. eXtended Reality (XR) smart glasses show promise for supporting people with CVI understand and interact with their environment, and early studies indicate strong user interest. However, most solutions remain confined to lab settings and are not ready for real-world use. This paper identifies 40 interrelated challenges to deploying XR smart glasses as assistive tools for CVI. These are organised into three tiers—Foundation, System, and Interface—across nine high-level domains. The challenges span technical, design, and evaluation gaps that must be addressed to move beyond prototypes. We call on the computer vision, HCI, and systems communities to treat accessibility-driven deployment as a core design goal, supported by cross-disciplinary collaboration and real-world evaluation.

1. Introduction

Cerebral Visual Impairment (CVI) is now the leading cause of vision loss in children in developed countries [26, 53, 63], and is expected to become a major cause of adult vision loss as these children grow up [8]. Unlike Ocular Vision Impairment (OVI), which is caused by problems with the eyes, CVI results from damage or delays in the brain's visual processing areas [62]. This often affects higher-level visual abilities, such as recognising objects and focusing attention, more than low-level functions like acuity or field of view [47, 58]. People with CVI also often have other neurological conditions, and their assistive technology (AT) needs are different from those with ocular vision loss [14].

Extended reality (XR) smart glasses have recently gained interest as an assistive tool for environment understanding and interaction due to their support for visual feedback, wearable, and hands-free nature [14, 16]. Early studies show that people with CVI are interested in using such devices in their daily lives to provide real-time support for tasks like reading, recognising objects, or identifying people [16]. However, most XR systems are still in early stages and work only in lab settings [32, 44]. There are many real-world challenges that must be solved before smart glasses can become reliable tools for everyday use.

This paper builds on two recent studies: a co-design study with people with CVI that explored their needs and prototyped smart glasses solutions [16], and a conceptual framework for wearable intelligent assistants developed through discussions with users, researchers, and developers [31]. Drawing on these studies and our own experience developing XR tools for CVI, we identify 40 key challenges in deploying XR smart glasses as assistive devices. Figure 1 shows how they are organised into nine high-level domains across three tiers: Foundation, System, and Interface. These tiers are interdependent—limitations at the foundational level often constrain system functionality and interface design.

Addressing these challenges is key to moving beyond prototypes toward deployable solutions that support people with CVI—and can also advance the broader field of assistive XR smart glasses. We invite researchers and developers to shift focus from short-term prototypes to long-term solutions that can be used in the real world.

2. Related Work

2.1. Smart Glasses and XR for Visual Assistance

Smart glasses, or head-mounted displays (HMDs), have gained traction as assistive technologies due to their wearable, hands-free design and support for visual tasks. Kim and Choi [34] reviewed 57 studies across domains such as surgery, industrial maintenance, and assistive support, finding that hands-free, real-time interaction was especially valuable in visually demanding, task-intensive settings.

For people with vision impairments, Li et al. [44] reviewed 41 studies focused on using HMDs for low vision support. These systems used various extended reality (XR) technologies—such as augmented, mixed, or virtual real-

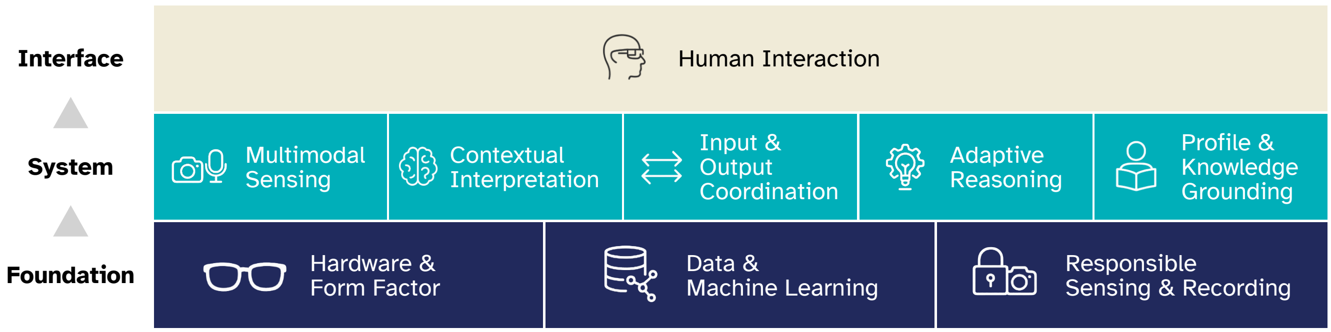

Figure 1. Three-tiered framework of deployment challenges for assistive XR smart glasses: The framework is organised into Foundation, System, and Interface levels, each with the high-level domains of the challenges.

ity—to improve perception and mobility. Augmented reality (AR) and mixed reality (MR) were found to be more useful for assistive purposes, while virtual reality (VR) was mostly used for therapy and training [18].

More recently, Kasowski et al. [32] reviewed 76 studies using XR technologies to support people with low vision. They described a range of techniques, including contrast enhancement [80, 81, 83], edge highlighting, and spatial audio cues to support navigation [29, 82]. While the findings show that XR tools can be helpful, the review also pointed out that most systems were not tested in real-world settings and lacked input from users during design. Importantly, none of these reviews found work for people with CVI, indicating a clear gap in research for brain-based visual impairments.

2.2. Assistive Technology for CVI

Research on AT for CVI has mainly focused on children, often through case studies [12, 39] or parent-reported experiences [21, 48]. There is very little work exploring how adults with CVI use AT in real-world settings. Gamage et al.[15] reviewed existing literature and found only three studies at the intersection of CVI and AT [7, 46, 59]. However, these studies mostly discussed ideas or adapted technologies originally designed for low vision, rather than providing solutions for CVI.

Through focus groups with people with CVI, Gamage et al. [14] identified seven key AT challenges: unawareness, locating, identifying, reading, sensory overload, mobility and luminance & contrast sensitivity. In a follow up study, they worked with two adults with CVI in an eight-month co-design process focused on developing XR smart glasses solutions [16]. The study found that smart glasses could support a range of daily activities, including locating objects, reading text, recognising people, engaging in conversations, and managing sensory stress (see Figure 2 for examples). Both participants expressed strong interest in using the device as part of their everyday lives.

However, the study also revealed several barriers to real-world deployment. Technical challenges such as environmental variability impacting model reliability and hardware limitations, and design issues like the need for hyper-personalisation and managing cognitive load, limited the system's performance outside of controlled environments.

2.3. The TOM Framework for Wearable Assistants

Janaka et al. [31] proposed a conceptual framework for building wearable intelligent assistants, called The Other Me (TOM). TOM is an open-source system that outlines the core components needed for developing context-aware wearable devices. It includes key components such as sensing the environment, understanding the user's context, and making decisions based on that information. While TOM is not specific to XR or assistive technology, it offers a useful way to think about the conceptual building blocks required when developing smart glasses.

In our work, we build on the co-design study, the TOM framework, and our own development experience to better frame the challenges of deploying XR smart glasses for people with CVI.

3. Approach

Our analysis started with the TOM framework [31], that broke down the system into 13 conceptual modules, such as context sensing, user sensing, coordination and reasoning. We then conducted an open-ended review to identify deployment challenges within each category. This drew on prior research on XR smart glasses for people with CVI, as well as broader work in low vision and assistive technologies where applicable. This process identified 75 challenges spanning technical, usability, and infrastructural issues.

The three authors then collaboratively consolidated overlapping challenges. For example, "understanding user" and "understanding context" were merged into a single domain: *Contextual Interpretation*. Similarly, recurring issues for example in data and machine learning across multiple areas were grouped under *Data and Machine Learning*. This process resulted in nine high-level domains.

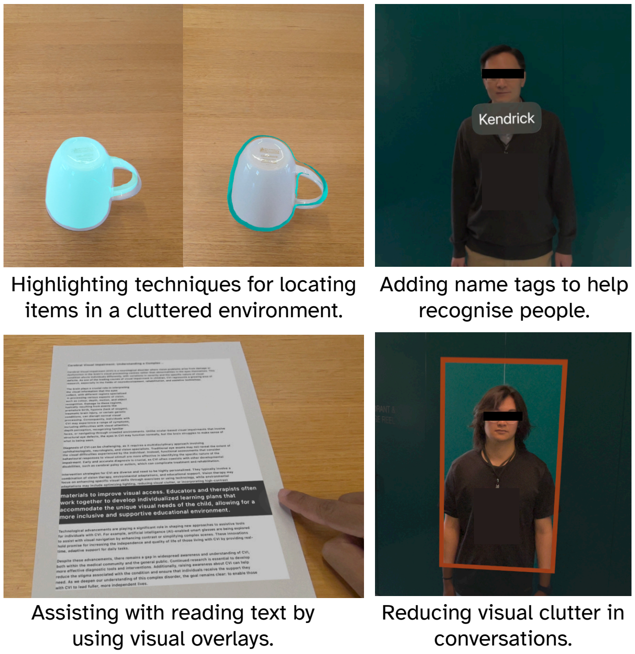

Figure 2. Examples of XR smart glasses solutions using the Apple Vision Pro for people with CVI, adapted from Gamage et al. [16].

Each challenge was then classified as either a Foundation, System, or Interface-level challenge (see Figure 1). The final result is a set of 40 deployment challenges organised across three tiers and nine high-level domains. The following sections (Sections 4, 5, and 6) present these challenges in detail.

4. Foundation-Level Challenges

These challenges stem from foundational barriers in hardware design, data and machine learning, and responsible sensing and recording.

4.1. Hardware and Form Factor

Smart glasses are constrained by hardware and form-factor challenges that cascade into every aspect of their design and performance [27, 56].

On-Device Compute Budgets: Wearables have limited processing power, making it difficult to run multiple models concurrently or maintain real-time responsiveness on demanding situations. Tasks such as object recognition, spatial mapping, and speech-to-text transcription compete for limited compute resources, often causing dropped frames, delayed feedback, or thermal throttling. Users also describe that system lag made them unsure whether their commands had been registered, or whether the system had failed silently [16]. While edge acceleration strategies can help [71], they must be balanced against accuracy and modality coverage, particularly in contexts where users rely on fast cues for orientation and safety.

Power and Battery: Bright AR displays, active depth sensors, and continuous audio feedback rapidly deplete the limited battery capacity typical of head-worn devices. Prior studies on AT report average runtime under two hours [27], which is likely to be insufficient for everyday tasks such as commuting, attending appointments, or navigating unfamiliar environments. Frequent charging interrupts the continuity of assistance and undermines user trust in the system's reliability. While energy-saving strategies like selectively disabling idle modules can extend runtime, they must be carefully aligned with user solutions to avoid suppressing timely cues and feedback.

Form-Factor Bulk and Thermal Constraints: Achieving a compact and comfortable form factor requires careful trade-offs between hardware features and physical design. While adding advanced sensors and processors can improve functionality, they also increase bulk and thermal output. Incorporating heat sinks or cooling components exacerbates this bulk, often resulting in eyewear that feels cumbersome or appears socially intrusive. Discomfort from device weight is a leading cause of prosthetic abandonment in prior AT research [56]. Developing lightweight, thermally efficient designs remains a critical yet unresolved engineering challenge for wearable smart glasses.

Display Modality Trade-Offs: XR smart glasses typically use either optical see-through (OST) or video see-through (VST) display architectures [40]. OST systems, such as Microsoft HoloLens, preserve direct view of the physical world through optics but often struggle with low-contrast overlays, especially in bright or outdoor environments. In contrast, VST systems like the Apple Vision Pro use passthrough camera feeds, enabling sharper, high-contrast augmentations but at the cost of mediating the user's access to the real scene. However, many commercial VST devices limit user control over passthrough processing, reducing their ability in passthrough manipulation [16]. Hybrid or adaptive display systems are still in early research stages and pose additional design and engineering challenges.

4.2. Data and Machine Learning

Recent advances in AI and machine learning have significantly expanded the capabilities of XR smart glasses [34]. However, delivering reliable experiences for people with CVI still depends on addressing these foundational challenges.

Data Scarcity and Label Quality: While numerous public datasets exist for general computer vision tasks [10, 45], they rarely capture scenarios specific to CVI such as environments that induce stress, biomarkers indicating cognitive load, or visual elements (e.g., color, contrast) that effectively capture attention. This limits the ability to train models that generalise to the needs of people with CVI.

Although prior work has highlighted the importance of accessibility-focused datasets [23], collecting such data is both resource-intensive and often hampered by inconsistent annotation standards. Synthetic data augmentation offers a partial solution, but typically falls short in capturing the nuanced contextual challenges experienced by people with CVI in real-world settings.

User Variability and Personalisation: CVI presents with high inter-individual variability, including challenges such as light streaks, difficulty in face recognition, and difficulty with object localisation [47, 58]. As a result, one-size-fits-all models are often ineffective, while highly personalised solutions face challenges related to data scarcity. Although techniques such as meta-learning and few-shot adaptation show promise for enabling rapid personalisation [78], they remain underdeveloped in real-world assistive contexts. Issues such as over-fitting [64] and catastrophic forgetting [9] continue to limit their practical deployment.

Model Inference Latency and Runtime: Even with mobile-optimised architectures such as MobileNetV3, inference on embedded hardware can introduce delays significant enough to disrupt the real-time responsiveness [24]. Such responsiveness is critical for tasks like obstacle avoidance or identifying shops from a moving vehicle for people with vision impairments. Attempts to reduce latency through techniques like model pruning or quantization often degrade prediction accuracy, highlighting a trade-off between speed and accuracy [37]. These are compounded when multiple models compete for limited compute, amplifying latency and reducing overall system responsiveness.

Local vs. Server Inference: Rich contextual reasoning demands extensive computational power, especially with large language models (LLMs). Off-loading inference to cloud servers supplies the necessary resources but introduces latency, connectivity dependence, and potential privacy issues. On-device inference, by contrast, provides low-latency, yet is limited by constrained compute budgets, often reducing model accuracy. Balancing these opposing constraints—fast but lightweight local models versus more capable yet slower and network-dependent server models—remains a key challenge.

Model Degradation: Over time, models may degrade due to changing environments (e.g., seasonal variation, construction zones) or shifts in user behaviour (e.g., after rehabilitation or training). For people who may rely on consistent performance across months or years, degradation without notification or remediation poses significant risks. While periodic retraining is the standard solution, it requires automatic data collection and labelling currently unavailable on wearables.

Continual Learning: Deployable systems must adapt to new objects, environments, and evolving user preferences without undergoing full retraining. Yet this process faces major challenges such as catastrophic forgetting, balancing stability and plasticity, and operating under resource constraints of wearable devices [73, 85]. Real-world data noise, shifting task distributions, and diverse usage scenarios add further complexity. Although numerous mitigation strategies exist [73], no single method fully addresses all challenges, highlighting the need for more robust and resource-efficient continual learning solutions.

LLM Hallucination and State Persistence: LLMs have shown potential as assistive agents by handling complex reasoning and dialogue tasks. However, beyond the engineering challenges of running LLMs efficiently on-device (see Model Inference and Runtime, Local vs. Server Inference), key limitations remain particularly in hallucination and poor state persistence. For example, an LLM might incorrectly detect an "EXIT" sign where none exists or forget critical user context, such as a peanut allergy, and falsely claim a product is safe—posing serious risks. While emerging research is exploring ways to ground LLMs using verified sensory inputs [69, 75, 84], these methods are rarely tailored to the unique needs of people with CVI.

4.3. Responsible Sensing and Recording

Continuous sensing is essential for smart glasses to provide real-time support, yet it introduces significant risks around privacy, data handling, and transparency.

Privacy Concerns: Always-on cameras and microphones raise ethical and legal issues, especially when recording occurs without clear user intent. Prior studies have shown that people with low vision often feel uneasy using devices with perceived surveillance tools [72]. Clear, accessible controls such as LEDs to signal recording are essential. Privacy must be a core design priority, not an afterthought, to ensure acceptance and wide adoption.

Edge-Case Logging with Contextual Annotation: Improving robustness for CVI-specific scenarios requires capturing rare but critical events such as sudden head or eye movements that are typically underrepresented in training data. However, recording these events along with contextual metadata (e.g., location, environment) raises significant privacy concerns. De-identifying and processing data closer with edge computing offers a promising direction. Still, there is a need to define operational, technical, and ethical standards for privacy-preserving logging and annotation in AT [5, 19].

Privacy-First Storage: Secure on-device capture is not enough if long-term storage lacks user control. Sensitive data such as banking details can still be exposed if the device is lost or compromised. Best practices include encrypted storage with user-defined retention and deletion settings, but ensuring transparent, tamper-proof deletion is critical for both user trust and regulatory compliance.

5. System-Level Challenges

These challenges are primarily engineering and implementation issues that hinder the deployment for real-world use.

5.1. Multimodal Sensing

Multimodal sensing enables smart glasses to perceive both the environment and the user, through explicit signals like hand gestures and implicit cues such as elevated heart rate during object search [31]. For people with CVI, who may experience changing vision and cognitive fatigue, relying on a single sensor modality is often inadequate [16]. Integrating multimodal sensor data enables more responsive and context-aware support, but also introduces new challenges in drift, reliability and protocols.

Sensing Accuracy and Drift: Sensors are prone to drift, misalignment, and noise, which can distort spatial cues and bio signals. This compromises the reliability of downstream systems that depend on accurate sensing. Studies in wearable AT have shown that sensor drift often goes unnoticed by users but can lead to sudden guidance failures that undermines trust [52, 55]. While progress has been made in mitigating these effects [11, 79], they remain a challenge for reliable use in real-world environments [55].

Environmental Robustness and Reliability: Sensor performance often degrades in real-world conditions such as low light, transitional lighting, or reflective surfaces. For example, reflections can distort depth sensing, and poor lighting can impair camera input. These issues are rarely captured in lab settings [16]. Assistive XR systems must be designed to handle such variability. At minimum, they should detect suboptimal sensor conditions and clearly communicate this to users for better awareness and trust.

Sensor Sampling Rate: Fixed sampling rates present trade-offs between accuracy and efficiency. Low-frequency sampling may miss critical physiological or contextual changes, while high-frequency sampling drains battery and generates heat. Adaptive sampling, which adjusts based on task or user state, has shown promise in wearables, but is still under explored in assistive XR applications [1, 3].

Biometric Data Standardisation: Inconsistent data standards across wearable platforms create major integration challenges. Biometric signals such as heart rate, gaze, and pupil dilation are often locked behind proprietary SDKs with incompatible formats and time-stamping. This fragmentation complicates real-time sensor access and increases the effort required for cross-platform development. Without standardised middleware protocols, hyper-personalised systems for people with CVI remain fragile and difficult to scale. While unified frameworks have gained traction in fields like healthcare and fitness [51], they are still largely absent in assistive technologies.

5.2. Contextual Interpretation

Accurately interpreting multimodal sensor data requires addressing challenges in aligning with context, such as user intent, task demands, and environmental conditions.

Concept Disambiguation: In dynamic environments, the classification of an object such as a person can vary depending on user intent, for example, whether they are considered an obstacle or a point of interaction. For instance, if a person with CVI is looking for the waiter at a cafe, misidentifying a waiter as an obstacle may lead the system to guide the user away instead of toward their intended interaction. Vision models trained on generic datasets often fail to capture this distinction, leading to misleading cues. Research shows that both visual context and prior knowledge influence how objects are interpreted in real time [35, 68], making it essential for systems to integrate both semantic understanding and situational awareness.

Contextual Framing: Recognising an object is only part of the challenge; systems must also understand its relevance to the user's current task. For example, detecting a person ahead could prompt different actions such as guiding the user to join a queue or warning them to stop at a crosswalk. While *Concept Disambiguation* focuses on correctly identifying the object ('person' or 'obstacle'), *Contextual Framing* determines the appropriate response based on that identification. One approach is to have users manually trigger tasks, but inferring intent from past behaviour and current actions offers a more seamless experience. Though it is technically complex, achieving this level of situational awareness remains an open challenge.

Temporal Context Alignment: Interpreting dynamic environments requires linking sensor data over time. For a person with CVI, a heart-rate spike while trying to locate a person in a visually cluttered street may signal stress that, if temporally linked, could prompt the system to simplify guidance. Without temporal alignment, the system treats events in isolation, missing opportunities to support the user more effectively. Mechanisms like rewindable logs [17] for models can help bridge these gaps, reducing cognitive load and improving system responsiveness.

Spatial and Interaction Continuity: Systems must preserve spatial coherence across visually complex environments. Many current solutions fail to maintain state through disruptions, such as when a tracked person is briefly occluded, forcing users to reselect targets. Unstable overlays, such as jittering or drifting arrows, can further disorient people with CVI [16]. These issues are especially problematic in cluttered or dynamic scenes, where occlusions disrupt both memory persistence and accurate world mapping. Advances in solutions like SLAM are critical to preserve interaction continuity and spatial alignment [22, 76].

5.3. Input and Output Coordination

Effective assistance depends on aligning multiple inputs, such as sensor data and contextual cues, with outputs delivered through XR overlays, audio, and haptic feedback, all while respecting computational constraints. However, several system-level challenges needs to be addressed.

Task Prioritisation Under Compute Constraints Real-time resource management is a key challenge when multiple tasks compete for limited compute. Non-critical processes must yield to more urgent tasks; for example, reading a sign should be de-prioritised if the user is actively navigating around obstacles. Static scheduling approaches often fail under load, causing latency. To maintain responsiveness, systems must adopt dynamic prioritisation that adjusts based on environmental context and user intent [61].

Multimodal Signal Coordination: Synchronising inputs like sensor data and contextual cues with outputs such as visual, audio, and haptic feedback requires real-time filtering and fusion to highlight relevant signals and reduce distractions. Even slight delays can disrupt the experience and divide the user's attention. For instance, if a user taps on text and the visual highlight appears before the audio cue, the mismatch can cause confusion. Reliable coordination depends on real-time synchronisation, automatic recovery from misalignment, and predictive buffering—all of which remain technically challenging in real-world conditions.

Notification Rate Control: Managing the frequency and timing of notifications is essential, particularly in visually cluttered environments. A high volume of notifications, such as bounding boxes appearing while scanning a crowded bookshelf, can overwhelm users and increase cognitive load. For people with CVI, this can lead to missed cues or task abandonment [16]. Adaptive systems that can batch, delay, or pace non-urgent prompts are important. Techniques like cooldown intervals and context-aware modulation, where notification rates adjust based on user activity (for example, walking versus standing), have shown promising results [25, 36].

Notification Modality Optimisation: Selecting the appropriate notification channel is critical for real-time assistance. Audio cues can be drowned out on busy streets, whereas haptic signals may feel intrusive in quiet settings. Effective delivery therefore requires sensing ambient conditions and adapting to user preferences. For language-focused tasks, people with CVI often prefer combined visual and audio feedback [16]. Systems must rapidly switch or blend modalities to convey essential information.

Abstraction Level Control: Delivering feedback at the proper level of detail is important. Commands that are too precise, such as "rotate 37°," can overwhelm users, while vague prompts like "turn right" are ambiguous in cluttered spaces. Visual guidance faces the same trade-off; in object-search tasks, people with CVI preferred an initial arrow for orientation followed by a highlight on the target object [16]. The optimal abstraction level should depend on the scene complexity. LLMs can help generate context-aware instructions, but further study is needed [28].

5.4. Adaptive Reasoning

Deployment of assistive XR smart glasses requires more than accurate perception. Systems must manage uncertainty, edge cases, and provide explanations for actions.

Conflict Resolution: Multimodal systems often face conflicting sensor inputs. For instance, the depth sensor might detect motion even when cameras sees a static scene. These discrepancies can lead to unsafe guidance if not properly managed. Fixed rules are too rigid and risk missing important context, while advanced solutions must weigh sensor confidence, past reliability, and the environment [4, 49].

Uncertainty Management: Ambiguity is a natural part of machine learning, and failing to communicate uncertainty can lead to serious risks. For instance, an error misclassifying a glass wall as an open path can put users in danger. To ensure safe use, systems should express uncertainty in ways that are clear but not distracting. Techniques like greyed-out overlays or prompts such as "Please verify visually" can help users recognise uncertain outputs [16]. Effectively communicating uncertainty supports safer decision-making and promotes collaboration between the user and system.

Explainability: Explainability involves two parts: understanding why the system made a decision and presenting that reasoning in a clear, usable form (often referred to as interpretability). Many models produce outputs without revealing the reasoning behind them, which limits user trust and understanding during real-time use. Techniques such as saliency maps, confidence scores, and attention visualisations show promise [67, 77], but they require alignment with both computational resources and user context. For example, if a system identifies a food item as gluten-free, it should explain whether this was based on a product label, a trusted database, or ingredient recognition. A simple label like "gluten-free" is not sufficient for users with medical needs. Clarifying messages such as "Identified from front label" or "Verified against certified database" allow users to assess reliability and decide if further checks are necessary. Clear, context-aware explanations are essential for safe and trustworthy assistive guidance.

Adaptation to Edge Cases: ML systems often struggle with rare but important edge cases, such as digital menu boards with changing layouts, mirrored elevators, or unconventional signage [57]. These scenarios are difficult to avoid in real-world environments, so handling them reliably is essential. This requires both model-level improvements (see *Data and Machine Learning*) and system-level responses, such as clear messages like "I'm unsure, proceed with caution" to help users navigate uncertainty safely.

5.5. Profile and Knowledge Grounding

Smart glasses must adapt to each user and their environment by maintaining up-to-date profiles of preferences, abilities, routines, and familiar spaces. The challenge is keeping this information current without introducing friction.

User Profile Acquisition Overhead: Personalisation depends on user-specific data, such as preferences, frequently visited locations, and familiar faces. However, traditional onboarding flows with long setup steps can feel burdensome and often lead to abandonment [13, 20, 66]. For example, [16] describes a people recognition feature that required manual entry of family members. A more intuitive approach would be to prompt the user after repeated encounters, such as "Would you like to name this person for future recognition?" The key challenge is designing methods for collecting meaningful profile data without disrupting the user experience.

Live Profile Adaptation: User preferences and abilities change over time due to therapy, aging, or shifts in environment. Static profiles can quickly become outdated, resulting in guidance that no longer fits the user's needs. For example, a person with CVI may gradually improve their ability to manage visual attention [16]. The challenge is to support dynamic profile updates by monitoring signals such as task patterns, performance, and user success.

Knowledge Base Synchronisation: Assistive features such as sign recognition, indoor navigation, and transit updates rely on accurate, up-to-date world knowledge. The core challenge is to synchronise these knowledge bases without interrupting or slowing down the user experience, particularly when dealing with large or frequently changing datasets. Efficient strategies like federated distillation, differential syncing and background updates during idle times can help while keeping content timely and relevant [86].

6. Interface-Level Challenges

These challenges focuses on XR smart glass specific interaction barriers that even with strong system-level performance can lead to abandonment and unsafe use.

6.1. Human Interaction

Many are classic HCI problems [2, 33, 41, 42, 50, 65], but their impact is amplified in assistive contexts.

Affordance and Interpretability: XR smart glasses introduce a unique affordance challenge: real-world objects often lack clear indicators of interactivity in AR environments [74]. Users may be unsure whether they can tap, select, or speak to an object, particularly when visual cues are subtle or inconsistent. This issue is especially challenging for people with CVI, who may struggle with low contrast, visual clutter, or ambiguous symbols. To enable intuitive interaction, objects should clearly convey their function through the use of consistent indicators. Addressing this challenge is critical to making XR systems not only interpretable but also usable in everyday assistive contexts.

Input Reliability and Accuracy: Multimodal inputs such as voice, gaze, and hand gestures often fail under real-world conditions. Background noise can disrupt speech recognition, gaze tracking may drift with attention shifts, and gestures are frequently misread or triggered unintentionally [2, 41, 50]. Hand-based input can be especially difficult for people with CVI, particularly those with limited fine motor control [16]. Identifying reliable input methods that align with users' specific abilities and constraints remains a core interaction challenge.

Latency and Responsiveness: Timely feedback is essential for maintaining situational awareness. Delays in visual overlays or audio prompts can interrupt user flow and lead to confusion. LLM-based systems often introduce high latency, with on-device inference taking over 30 seconds and cloud responses up to 10 seconds [43]. The challenge lies not only in reducing these delays (see Section 4: *Data and Machine Learning*), but also in communicating latency and maintaining a smooth, reliable user experience.

Ergonomics and Accessibility: Smart glasses often require precise hand gestures, which can be difficult for people with motor or coordination impairments. These demands may cause fatigue, accidental inputs, or discomfort, adding to the input reliability challenges discussed earlier. Hardware design also affects usability; heavy frames, unbalanced batteries, or poorly placed sensors can cause strain and reduce wearability [42, 56]. Ergonomic and accessible design—both in hardware and interaction—is crucial for long-term comfort and inclusive adoption.

Learnability: XR smart glasses often combine multiple input modes (gaze, voice, gesture and app-based controls), each with distinct interaction styles that can overwhelm or frustrate users. Given the affordance challenges, clear guidance and gradual on-boarding are essential. Simplified setup, consistent feedback, and staged feature introduction help build user confidence and support long-term use, but remain a key design challenge [30].

Feedback and Confirmation: Immediate, clear feedback helps users confirm their input has been received and understood [65]. Without cues like audio signals or visual indicators, users may become uncertain, leading to repeated actions or hesitation [16]. The challenge is to provide timely feedback while managing the output challenges discussed in *Input and Output Coordination*.

7. Discussion

This paper outlines 40 interrelated challenges spanning three tiers that are barriers for real-world deployment of XR smart glasses for people with CVI. This tiered structure reveals that deployment is not just a matter of technical readiness but also of aligning system behaviour and interaction design with real-world needs.

A key insight from this three-tier lens is the need to balance bottom-up and top-down strategies across these tiers. A bottom-up strategy focuses on improving core technologies such as sensing, data quality, and latency, but may delay user-facing progress. A top-down strategy, starting from lived experiences and daily needs, can help prioritise which technical improvements matter most. We propose that real-world deployment requires a dual approach: aligning technical development with user context and grounding user solutions in technical feasibility.

Some of these issues are starting to gain research attention, including explainability [77], responsible data practices [19], and display modality trade-offs [40]. Others, such as XR-specific affordances, biometric data standards, and notification management, are still underexplored.

However, real-world deployment is also shaped by broader constraints beyond the three tiers, such as **Cost and Funding**. Development and real-world deployment of XR smart glasses requires sustained investment. The cost of hardware, data collection and user involvement can be a major barrier—especially if intended for everyday use. Partnering with commercial stakeholders and public institutions can help distribute costs and accelerate development.

7.1. Broader Implications

Tackling these challenges calls for coordinated progress across technical, interaction, and systems-level domains and opens up both short-term and long-term opportunities for research.

Computer Vision: Existing models are typically trained in controlled settings and evaluated on fixed benchmarks. Real-world deployment requires models that are context-aware, robust to ambiguity, and capable of adapting in dynamic, noisy environments—particularly on edge devices.

Sensor Fusion & Embedded Systems: Coordinating data from asynchronous, noisy sources such as cameras, depth sensors, IMUs, and microphones remains difficult under constraints of mobility, latency, and power. Research must develop fusion techniques that are efficient and reliable on wearable hardware.

Human–Computer Interaction (HCI): Designing for CVI challenges introduce new design directions. Interfaces must account for cognitive load, perceptual variability, and alternative feedback preferences. Inclusive co-design and adaptive interface strategies are essential.

Multimodal AI: There is a growing need for systems that integrate visual, auditory, and tactile feedback based on real-time context and user state. This shifts the focus from static, single-modality outputs to dynamic, context-sensitive interaction strategies.

ML Systems and Infrastructure: Reliable deployment requires infrastructure that supports continual learning, privacy-preserving adaptation, and on-device inference. This includes developing pipelines that can operate with sparse labels, noisy inputs, and intermittent connectivity.

Short-term opportunities include:
- Enabling access to eye tracking data through manufacturer APIs to support adaptive feedback and attention-aware interaction.
- Developing benchmark datasets that capture CVI-relevant conditions such as clutter, motion sensitivity, and luminance variability.
- Prototyping task-aware XR guidance that adapts to intent, such as distinguishing between exploration and navigation modes.
- Developing design toolkits to help researchers model CVI-relevant constraints.

Longer-term directions include:
- Building vision models that gracefully degrade or offer fallback strategies under uncertainty.
- Creating wearable platforms that adapt sensing and feedback based on user fatigue or cognitive state.
- Establishing shared standards for ethical, transparent, and user-controlled data handling in wearables.

8. Call to Action

While investment in XR technologies continues to grow, most commercial systems remain focused on entertainment, productivity, or social media—not accessibility [54]. This gap leaves people with vision impairments underserved and reinforces the marginal status of AT as an afterthought.

Some of today's most widely adopted features originated from work to support accessibility and were later generalised for broader use [6, 38, 60, 70]. For instance, speech-to-text systems were initially developed to support people with hearing loss, but now power mainstream products like live captions, assistants, and transcription tools [38, 70].

We call on the computer vision, HCI, and systems communities to treat deployment in accessibility as a first-class design goal, not a downstream application. This shift requires collaborative partnerships, new evaluation paradigms, and a commitment to real-world complexity.

9. Conclusion

XR smart glasses hold real promise as assistive technology for people with CVI, but realising this potential requires addressing critical deployment challenges. Spanning three tiers, the 40 challenges outlined in this paper offers a roadmap for future progress. While we do not propose technical implementations or quantitative validation, our goal is to spark cross-disciplinary collaboration and highlight how designing for accessibility can drive innovation toward more human-centered XR systems.

References

[1] O. Amft, F. Cruciani, Giovanni Schiboni, Juan Carlos Suarez, and Celia Martín Vicario. Context-adaptive sub-nyquist sampling for low-power wearable sensing systems. *IEEE Transactions on Mobile Computing*, 21:4249–4262, 2022. 5

[2] P. Arpaia, M. Parvis, Antonio Esposito, and N. Moccaldi. A single-channel and non-invasive wearable brain-computer interface for industry and healthcare. *Journal of Visualized Experiments : JoVE*, 197, 2023. 7

[3] N. S. Artan and R. Castro. Adaptive sampling for low-power wearable and implantable devices. *2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems Workshops (MASSW)*, pages 63–66, 2019. 5

[4] Christson Awanyo and Nawal Guermouche. Attention-driven conflict management in smart IoT-based systems. In *International Conference on Service-Oriented Computing*, pages 133–141. Springer, 2024. 6

[5] Daniel Biedermann, George-Petru Ciordas-Hertel, H. Drachsler, Julia Mordel, and Marc Winter. Contextualized logging of on-task and off-task behaviours during learning. *Journal of Learning Analytics*, 10:115–125, 2023. 4

[6] Jeffrey P. Bigham and Chris Fleizach. System-class accessibility. *Queue*, 22:28–39, 2024. 8

[7] Faith A. Birnbaum, Steven A. Hackley, and Lenworth N. Johnson. Enhancing visual performance in individuals with cortical visual impairment (homonymous hemianopsia): Tapping into blindsight. *Journal of Medical Hypotheses and Ideas*, 9(2):S8–S13, 2015. 2

[8] Daniëlle G. M. Bosch, F. Nienke Boonstra, Michèl A. A. P. Willemsen, Frans P. M. Cremers, and Bert de Vries. Low vision due to cerebral visual impairment: differentiating between acquired and genetic causes. *BMC Ophthalmology*, 14 (1):1–9, 2014. 1

[9] Mengjun Cheng, Hanli Wang, and Yu Long. Meta-learning-based incremental few-shot object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 32: 2158–2169, 2022. 4

[10] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255. Ieee, 2009. 3

[11] Tadashi Fujisawa, S. Tadano, H. Tohyama, Ryo Takeda, L. Gastaldi, and G. Lisco. Drift removal for improving the accuracy of gait parameters using wearable sensor systems. *Sensors (Basel, Switzerland)*, 14:23230 – 23247, 2014. 5

[12] Melody Zagami Furze and John P Phillips. Integrating functional MRI information into the educational plan of a child with cerebral visual impairment: A case study. *Journal of Visual Impairment & Blindness*, 112(5):532–540, 2018. 2

[13] Bhanuka Gamage, Thanh-Toan Do, Nicholas Seow Chiang Price, Arthur Lowery, and Kim Marriott. What do blind and low-vision people really want from assistive smart devices? comparison of the literature with a focus study. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*, New York, NY, USA, 2023. Association for Computing Machinery. 7

[14] Bhanuka Gamage, Leona Holloway, Nicola McDowell, Thanh-Toan Do, Nicholas Price, Arthur Lowery, and Kim Marriott. Vision-based assistive technologies for people with cerebral visual impairment: A review and focus study. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*, New York, NY, USA, 2024. Association for Computing Machinery. 1, 2

[15] Bhanuka Gamage, Leona Holloway, Nicola McDowell, Thanh-Toan Do, Nicholas Seow Chiang Price, Arthur James Lowery, and Kim Marriott. Broadening our view: Assistive technology for cerebral visual impairment. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–9, 2024. 2

[16] Bhanuka Gamage, Nicola McDowell, Dijana Kovacic, Leona Holloway, Thanh-Toan Do, Nicholas Price, Arthur Lowery, and Kim Marriott. Smart glasses for cvi: Co-designing extended reality solutions to support environmental perception by people with cerebral visual impairment. *arXiv preprint arXiv:2506.19210*, 2025. 1, 2, 3, 5, 6, 7

[17] Rolando Garcia, Eric Liu, Vikram Sreekanti, Bobby Yan, Anusha Dandamudi, Joseph E. Gonzalez, Joseph M. Hellerstein, and Koushik Sen. Hindsight logging for model training. *arXiv preprint arXiv:2006.07357*, 2020. 5

[18] Franziska Geringswald, Eleonora Porracin, and Stefan Pollmann. Impairment of visual memory for objects in natural scenes by simulated central scotomata. *Journal of Vision*, 16 (2):6–6, 2016. 2

[19] Sahra Ghalebikesabi, Eugene Bagdasaryan, Ren Yi, Itay Yona, Ilia Shumailov, Aneesh Pappu, Chongyang Shi, Laura Weidinger, Robert Stanforth, Leonard Berrada, et al. Operationalizing contextual integrity in privacy-conscious assistants. *arXiv preprint arXiv:2408.02373*, 2024. 4, 8

[20] Eleni Gkiolnta, Debopriyo Roy, and G. Fragulis. Challenges and ethical considerations in implementing assistive technologies in healthcare. *Technologies*, 2025. 7

[21] Trudy Goodenough, Anna Pease, and Cathy Williams. Bridging the gap: Parent and child perspectives of living with cerebral visual impairments. *Frontiers in Human Neuroscience*, 15:689683, 2021. 2

[22] Junlong Guo, Hesheng Yin, Shaomiao Li, Yu Tao, and Bo Huang. Dynam-SLAM: An accurate, robust stereo visual-inertial SLAM method in dynamic environments. *IEEE Transactions on Robotics*, 39:289–308, 2023. 5

[23] Danna Gurari, Qing Li, Abigale J. Stangl, Anhong Guo, Chi Lin, Kristen Grauman, Jiebo Luo, and Jeffrey P. Bigham. Vizwiz grand challenge: Answering visual questions from blind people. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3608–3617, 2018. 4

[24] Jin Han and Y. Yang. Real-time object detector based MobileNetV3 for UAV applications. *Multimedia Tools and Applications*, 82:18709–18725, 2022. 4

[25] C. Hargood, V. Pejović, Mirco Musolesi, Scott Lloyd, L. Morrison, L. Yardley, Natalie Goodman, M. Weal, A. Weston, Adam W. A. Geraghty, and D. Michaelides. The effect of timing and frequency of push notifications on usage of a smartphone-based stress management intervention: An exploratory trial. *PLoS ONE*, 12, 2017. 6

[26] Deborah D Hatton, Eric Schwietz, Burt Boyer, and Paul Rychwalski. Babies count: The national registry for children with visual impairments, birth to 3 years. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, 11(4):351–355, 2007. 1

[27] Marion Hersh. Wearable travel aids for blind and partially sighted people: A review with a focus on design issues. *Sensors*, 22(14):5454, 2022. 3

[28] Heidi Ahmed Holiel, S. Fawzi, and Walid Al-Atabany. Assisting visually impaired subjects using large language models: A comprehensive evaluation. *2024 6th Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, pages 561–566, 2024. 6

[29] Katsuya Hommaru and Jiro Tanaka. Walking support for visually impaired using AR/MR and virtual braille block. In *International Conference on Human-Computer Interaction*, pages 336–354. Springer, 2020. 2

[30] Magdalena Igras-Cybulska, Mateusz Danioł, Magdalena Wójcik-Pedziwiatr, Daria Hemmerling, Jakub Kaminski, Pawel Jemiolo, and Marek Wodziński. Designing accessible XR for neurodegenerative disease patients: Insights from Parkinson's disease case study. *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 1174–1175, 2025. 7

[31] Nuwan Janaka, Shengdong Zhao, David Hsu, Sherisse Tan Jing Wen, and Chun Keat Koh. Tom: A development platform for wearable intelligent assistants. In *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*, page 837–843, New York, NY, USA, 2024. Association for Computing Machinery. 1, 2, 5

[32] Justin Kasowski, Byron A. Johnson, Ryan Neydavood, Anvitha Akkaraju, and Michael Beyeler. A systematic review of extended reality (xr) for understanding and augmenting vision loss. *Journal of Vision*, 23(5):5–5, 2023. 1, 2

[33] S. Kennison, Aaron Cecil-Xavier, Alireza Sadeghi Milani, J. Cecil, and Avinash Gupta. A systematic review of human–computer interaction (HCI) research in medical and other engineering fields. *International Journal of Human–Computer Interaction*, 40:515 – 536, 2022. 7

[34] Dawon Kim and Yosoon Choi. Applications of smart glasses in applied sciences: A systematic review. *Applied Sciences*, 11:4956, 2021. 1, 3

[35] P. Knoeferle, M. Pickering, M. Crocker, and Christoph Scheepers. The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, 95:95–127, 2005. 5

[36] B. Kröse, Shihan Wang, H. V. Hoof, and Chao Zhang. Optimizing adaptive notifications in mobile health interventions systems: Reinforcement learning from a data-driven behavioral simulator. *Journal of Medical Systems*, 45, 2021. 6

[37] Uday Kulkarni, Meena S. M., Sunil V. Gurlahosur, and Gopal Bhogar. Quantization friendly MobileNet (QF-MobileNet) architecture for vision based applications on embedded platforms. *Neural networks : The Official Journal of the International Neural Network Society*, 136:28–39, 2020. 4

[38] Y. Kumar, Chamkaur Singh, and Apeksha Koul. A deep learning approaches in text-to-speech system: A systematic review and recent research perspective. *Multimedia Tools and Applications*, 82:15171–15197, 2022. 8

[39] Katie Lane-Karnas. A case study on cerebral visual impairment, reading, and braille. *Journal of Visual Impairment & Blindness*, 117(6):498–504, 2023. 2

[40] Tobias Langlotz, Jonathan Sutton, and Holger Regenbrecht. A design space for vision augmentations and augmented human perception using digital eyewear. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–16, 2024. 3, 8

[41] Lik-Hang Lee and Pan Hui. Interaction methods for smart glasses: A survey. *IEEE Access*, 6:28712–28732, 2017. 7

[42] G. Leong, Abderahman Rejeb, John G. Keogh, and Horst Treiblmaier. Potentials and challenges of augmented reality smart glasses in logistics and supply chain management: A systematic literature review. *International Journal of Production Research*, 59:3747 – 3776, 2021. 7

[43] Luchang Li, Sheng Qian, Jie Lu, Lunxi Yuan, Rui Wang, and Qin Xie. Transformer-lite: High-efficiency deployment of large language models on mobile phone gpus. *arXiv preprint arXiv:2403.20041*, 2024. 7

[44] Yifan Li, Kangsoo Kim, Austin Erickson, Nahal Norouzi, Jonathan Jules, Gerd Bruder, and Gregory F. Welch. A scoping review of assistance and therapy with head-mounted displays for people who are visually impaired. *ACM Transactions on Accessible Computing (TACCESS)*, 2022. 1

[45] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference*, pages 740–755. Springer, 2014. 3

[46] Marie-Céline Lorenzini and Walter Wittich. Factors related to the use of magnifying low vision aids: A scoping review. *Disability and Rehabilitation*, 42(24):3525–3537, 2020. 2

[47] Amanda H. Lueck, Gordon N. Dutton, and Sylvie Chokron. Profiling children with cerebral visual impairment using multiple methods of assessment to aid in differential diagnosis. In *Seminars in Pediatric Neurology*, pages 5–14. Elsevier, 2019. 1, 4

[48] Marta Lupón, Manuel Armayones, and Genís Cardona. Quality of life among parents of children with visual impairment: A literature review. *Research in Developmental Disabilities*, 83:120–131, 2018. 2

[49] Sri Harish Mallidi, Roland Maas, Kyle Goehner, Ariya Rastrow, Spyros Matsoukas, and Björn Hoffmeister. Device-directed utterance detection. *arXiv preprint arXiv:1808.02504*, 2018. 6

[50] Gilda Manfredi, Pietro Carratu, N. Capece, and Vincenzo Macellaro. An easy hand gesture recognition system for xr-based collaborative purposes. *2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*, pages 121–126, 2022. 7

[51] T. Manini, Amal A. Wanigatunga, Parisa Rashidi, Sanjay P. Nair, Tonatiuh V Mendoza, A. Davoudi, D. Corbett, Sanjay

Ranka, and Matin Kheirkhahan. A smartwatch-based framework for real-time and online assessment and mobility monitoring. *Journal of Biomedical Informatics*, 89:29–40, 2019. 5

[52] Uriel Martinez-Hernandez, L. Jabban, Dingguo Zhang, B. Metcalfe, James Male, and T. Assaf. Wearable assistive robotics: A perspective on current challenges and future trends. *Sensors*, 21, 2021. 5

[53] Carey A. Matsuba and James E. Jan. Long-term outcome of children with cortical visual impairment. *Developmental Medicine and Child Neurology*, 48(6):508–512, 2006. 1

[54] Stephanie M. Mello, Jessyca L. Derby, Nicholas R. Horn, Barbara S. Chaparro, and Andi N. StClair. Examining the inclusivity of extended reality (xr) in current products. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 67:2198 – 2204, 2023. 8

[55] Farhaan Mirza, M. Baig, H. Gholamhosseini, and Shereen Afifi. A systematic review of wearable sensors and IoT-based monitoring applications for older adults – a focus on ageing population and independent living. *Journal of Medical Systems*, 43:1–11, 2019. 5

[56] Chantelle Neal, T. Packham, C. Wilkins, and Lauren C. Smail. Comfort and function remain key factors in upper limb prosthetic abandonment: Findings of a scoping review. *Disability and Rehabilitation: Assistive Technology*, 16:821 – 830, 2020. 3, 7

[57] Alice Othmani, Yahaya Idris Abubakar, A. Q. M. Sabri, and Patrick Siarry. A systematic review of rare events detection across modalities using machine learning and deep learning. *IEEE Access*, 12:47091–47109, 2024. 6

[58] Swetha Sara Philip and Gordon N. Dutton. Identifying and characterising cerebral visual impairment in children: A review. *Clinical and Experimental Optometry*, 97(3):196–208, 2014. 1, 4

[59] Kevin M. Pitt and John W McCarthy. Strategies for highlighting items within visual scene displays to support augmentative and alternative communication access for those with physical impairments. *Disability and Rehabilitation: Assistive Technology*, 18(8):1319–1329, 2023. 2

[60] Nirvana Popescu and Valentin Bercaru. A systematic review of accessibility techniques for online platforms: Current trends and challenges. *Applied Sciences*, 2024. 8

[61] Paul B. Reverdy and D. Koditschek. A dynamical system for prioritizing and coordinating motivations. *SIAM Journal on Applied Dynamical Systems*, 17:1683–1715, 2017. 6

[62] Hanna E. A. Sakki, Naomi J. Dale, Jenefer Sargent, Teresa Perez-Roche, and Richard Bowman. Is there consensus in defining childhood cerebral visual impairment? a systematic review of terminology and definitions. *British Journal of Ophthalmology*, 102(4):424–432, 2018. 1

[63] Lisbeth Sandfeld Nielsen, Liselotte Skov, and Hanne Jensen. Visual dysfunctions and ocular disorders in children with developmental delay. I. prevalence, diagnoses and aetiology of visual impairment. *Acta Ophthalmologica Scandinavica*, 85 (2):149–156, 2007. 1

[64] B. Schiele, Yaoyao Liu, Qianru Sun, and Tat-Seng Chua. Meta-transfer learning for few-shot learning. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 403–412, 2018. 4

[65] Ben Shneiderman and Catherine Plaisant. *Designing the user interface: Strategies for effective human-computer interaction*. Pearson Education India, 2010. 7

[66] M. Taugher, Roger O. Smith, J. Lenker, and F. Harris. Consumer perspectives on assistive technology outcomes. *Disability and Rehabilitation: Assistive Technology*, 8:373 – 380, 2013. 7

[67] Erico Tjoa, Cuntai Guan, Tushar Chouhan, and Hong Jing Khok. Enhancing the confidence of deep learning classifiers via interpretable saliency maps. *Neurocomputing*, 562: 126825, 2023. 6

[68] S. Tobimatsu, K. Ogata, Y. Kume, Takahiro Kimura, and T. Urakawa. Temporal dynamics of the knowledge-mediated visual disambiguation process in humans: A magnetoencephalography study. *European Journal of Neuroscience*, 41, 2015. 5

[69] Peter Tong, Ellis Brown, Penghao Wu, Sanghyun Woo, Adithya Jairam Vedagiri IYER, Sai Charitha Akula, Shusheng Yang, Jihan Yang, Manoj Middepogu, Ziteng Wang, et al. Cambrian-1: A fully open, vision-centric exploration of multimodal LLMs. *Advances in Neural Information Processing Systems*, 37:87310–87356, 2024. 4

[70] Ayushi Trivedi, Navya Pant, Pinal Shah, Simran Sonik, and Supriya Agrawal. Speech to text and text to speech recognition systems-areview. *IOSR Journal of Computer Engineering*, 20(2):36–43, 2018. 8

[71] Yigit Tuncel, Anish Krishnakumar, Aishwarya Lekshmi Chithra, Younghyun Kim, and Umit Ogras. A domain-specific system-on-chip design for energy efficient wearable edge AI applications. In *Proceedings of the ACM/IEEE International Symposium on Low Power Electronics and Design*, New York, NY, USA, 2022. Association for Computing Machinery. 3

[72] Andrew Utt, Allison J. Chen, Brian J. Nguyen, Daniel L. Chao, Ryan Apgar, William S. Chen, and Emily Hill. Large-scale assessment of needs in low vision individuals using the Aira assistive technology. *Clinical Ophthalmology*, 13:1853 – 1868, 2019. 4

[73] Buddhi Wickramasinghe, Kaushik Roy, and Gobinda Saha. Continual learning: A review of techniques, challenges, and future directions. *IEEE Transactions on Artificial Intelligence*, 5:2526–2546, 2024. 4

[74] Marco Winckler, D. Trevisan, E. Oliveira, E. Clua, Victor Vieira, and Aline Menin. Understanding affordances in XR interactions through a design space. *Proceedings of the XXIII Brazilian Symposium on Human Factors in Computing Systems*, 2024. 7

[75] Qirui Yang, Huatao Xu, Mani B. Srivastava, Mo Li, and Liying Han. Penetrative AI: Making LLMs comprehend the physical world. *Proceedings of the 25th International Workshop on Mobile Computing Systems and Applications*, 2023. 4

[76] Hongpei Yin, Sai Manoj Prakhya, Haojie Bai, Xiongwei Zhao, Yang Wang, Congcong Wen, Yijiao Sun, Jie Xu, and Run Zhou. Multimodal features and accurate place

recognition with robust optimization for lidar–visual–inertial SLAM. *IEEE Transactions on Instrumentation and Measurement*, 73:1–16, 2024. 5

[77] Jindi Zhang, Antoni B. Chan, Guoyang Liu, and J. Hsiao. Human attention-guided explainable artificial intelligence for computer vision models. *Neural Networks : The Official Journal of the International Neural Network Society*, 177: 106392, 2023. 6, 8

[78] Tengfei Zhang, Zhengyuan Zhang, M. Yan, Yue Zhang, Xian Sun, Z. Chang, and Kun Fu. Meta-SSD: Towards fast adaptation for few-shot object detection with meta-learning. *IEEE Access*, 7:77597–77606, 2019. 4

[79] Hongyu Zhao, S. Qiu, Huosheng Hu, and Zhelong Wang. Using distributed wearable sensors to measure and evaluate human lower limb motions. *IEEE Transactions on Instrumentation and Measurement*, 65:939–950, 2016. 5

[80] Yuhang Zhao, Sarit Szpiro, and Shiri Azenkot. Foresee: A customizable head-mounted vision enhancement system for people with low vision. In *Proceedings of the 17th international ACM SIGACCESS Conference on Computers & Accessibility*, pages 239–249, 2015. 2

[81] Yuhang Zhao, Sarit Szpiro, Jonathan Knighten, and Shiri Azenkot. Cuesee: Exploring visual cues for people with low vision to facilitate a visual search task. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 73–84, 2016. 2

[82] Yuhang Zhao, Elizabeth Kupferstein, Brenda Veronica Castro, Steven Feiner, and Shiri Azenkot. Designing AR visualizations to facilitate stair navigation for people with low vision. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pages 387–402, 2019. 2

[83] Yuhang Zhao, Elizabeth Kupferstein, Hathaitorn Rojnirun, Leah Findlater, and Shiri Azenkot. The effectiveness of visual and audio wayfinding guidance on smartglasses for people with low vision. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2020. 2

[84] Yang Zhao, Zhijie Lin, Daquan Zhou, Zilong Huang, Jiashi Feng, and Bingyi Kang. Bubogpt: Enabling visual grounding in multi-modal LLMs. *arXiv preprint arXiv:2307.08581*, 2023. 4

[85] Jun Zhu, Xingxing Zhang, Liyuan Wang, and Hang Su. A comprehensive survey of continual learning: Theory, method and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46:5362–5383, 2023. 4

[86] Kun Zhu, Xiaolan Lu, Juan Li, and Yang Zhang. Efficient knowledge base synchronization in semantic communication network: A federated distillation approach. *2024 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6, 2024. 7